

12

COMPUTER-AIDED EDITING - PRESENT AND FUTURE

Dr. J. Douglas Kniffin, Westinghouse Electric Corporation

This paper describes computer-aided editing systems which analyze English text and suggest improvements. The systems described are vastly different from the familiar word processing programs which permit editing changes such as insertions and deletions of characters, words, or lines of text. These new systems analyze the style and style difficulty of text and provide feedback to writers and editors for improving readability and comprehensibility. Three systems designed originally for examining the complexity of technical documents are described. These are the Bell Laboratories Writer's Workbench, the U.S. Navy Computer Readability Editing System, and the Westinghouse Writing Aids System. Though in varying stages of development, the features of the three systems are listed and compared and the research or other basis behind the editing features is discussed. Some limited data on the use of the systems and their acceptance by users are also presented. The paper concludes with a discussion of future applications of the systems such as writing instruction, style comparison, and computer-aided translation.

1. INTRODUCTION

For thousands of years, writing and editing were unmechanized processes. The printing press and the typewriter improved the legibility of documents but were of little help in the creative writing process. To produce text that was readable, useable, grammatically correct, well organized, and suitable for the intended audience, writers and editors relied on a variety of traditional aids. These aids included personal

knowledge, dictionaries, grammar books, and style guides. Advice was plentiful but often hard to follow consistently because of deadlines, drudgery, and human limitations. In this century, growth in the economy, government, and technology led to a tremendous increase in the need for documentation. The need for more documentation led to the need to automate the writing process.

Computer-based word processors began to appear in the early 1970's. The

first word processors did not aid the creative writing process, but they helped to speed up the production of documentation. Once text was captured in machine-readable form, it could be manipulated in many ways. Writing and editing changes were manually marked on a word processor printout, but the changes could then be keyboarded for automatic insertion into a document. Features such as "global search and replace" made it possible to make numerous corrections in a single operation. The need for retyping a changed manuscript was eliminated, and the final version of a document could be reformatted as desired. The early word processors were a great help in document production, but more automation was needed.

By the mid 1970's, researchers began to realize that the full computational powers of word processors were not being used. For example, if a unique character string could be recognized, why not recognize all unique character strings? In other words, why not program the word processor to check spelling? Similarly, if any ASCII character could be recognized, why not use this power to count words, sentences, and the number of characters or syllables in a word? In other words, why not program the word processor to calculate the readability grade level of text?

Programs for spelling checks and readability measurements quickly became the first automated aids to the creative writing processes. The programs were faster and generally more accurate than humans; they eliminated a great deal of boredom and drudgery. The time formerly used to check spelling could be devoted to other efforts. A readability analysis of a manuscript helped writers and editors judge its appropriateness to the intended audience. The success of spelling check and readability programs

sparked interest in developing more elaborate automated aids for writers and editors.

Independently and nearly simultaneously, researchers at Bell Laboratories, the U.S. Navy's Training Analysis and Evaluation Group (TAEG), and the Integrated Logistics Support Division of Westinghouse conceived the same basic idea. Why not use the computer or word processor to automatically analyze text for compliance with clear writing principles? Automatic text analysis could provide feedback on writing style, style difficulty, vocabulary, diction, etc. Text analysis could also provide suggestions for improving the grammar, punctuation, readability, and comprehensibility of documents. Freed of many time consuming and tedious manual tasks, writers and editors could concentrate on other considerations in document design such as organization and content. With so many obvious benefits to be gained, research on automatic text analysis began in earnest.

The early research at Bell Laboratories was directed at style analysis and identifying faulty diction; TAEG concentrated first on style analysis and vocabulary control; and Westinghouse worked first on style analysis and research-proven suggestions for text improvement. The research, though not complete, has led to the development of Bell Laboratories' Writer's Workbench programs, TAEG's Computer Readability Editing System (CRES), and Westinghouse's Writing Aids System (WRITEAIDS). (Other organizations including IBM and The University of Michigan are also working on automatic text analysis.)

Section 2 of this paper will discuss the major features of the Writer's Workbench, CRES, and WRITEAIDS which have been automated or are under development. Section 3 will

describe the feedback provided by the systems. Section 4 will discuss some limited data on user acceptance. Section 5 will address future applications of the systems.

2. FEATURES OF THE WRITER'S WORKBENCH, CRES, AND WRITEAIDS

The text analysis features of the Writer's Workbench, CRES, and WRITEAIDS now vary considerably even though the ultimate purpose of the systems is the same. However, as research progresses, the three systems are becoming more and more similar. The Writer's Workbench programs run under the UNIX* operating system. Originally, the CRES programs were written for computers or word processors with fairly large memory capacities. The CRES programs are now being rewritten for smaller stand-alone word processors. The WRITEAIDS programs were designed from the start for small stand-alone word processors.

Text analysis features tend to fall into three main categories although there is clearly some overlap in the categories. In the first category are straightforward features which are loosely defined here as customs of English language use. The second category, suggestions from writing experts, includes some "grammar rules" and clear writing principles. The third category, suggestions based on research, includes clear writing principles based on research in readability, psychology, linguistics, etc. The paragraphs which follow discuss the similarities and differences among the three systems and will discuss the rationale for the features where appropriate.

2.1 STRAIGHTFORWARD FEATURES

Table 1 summarizes the major straightforward features. Bell

Laboratories has implemented all of the features in this category. TAEG and Westinghouse have not yet emphasized research in this area to any great extent.

2.1.1 Spelling Check

The Writer's Workbench spelling check program has an internally developed 30,000 word dictionary with affixes.¹ The program also allows users to add words to tailor the program for themselves. Interactive program features allow users to locate possible misspellings, determine correct spelling, change spelling if necessary, and add words to a personal dictionary.²

CRES provides a spelling check in an indirect but interesting way. Any word that is not on the Master Word List is flagged as an uncommon word.³ The flagged words are thus possible misspellings.

WRITEAIDS uses commercially available spelling check programs.

2.1.2 Punctuation Check

A Writer's Workbench program searches for simple punctuation errors. When this program finds an error it prints the original line and a corrected line. It recommends changes to (1) capitalize the first letter of a sentence, (2) move commas and periods to the left of double quotes and move semicolons and colons to the right of double quotes, and (3) balance double or single quotes and parenthesis.⁴

Punctuation checks are not planned for CRES or WRITEAIDS.

2.1.3 Double Word Check

When the same word appears twice in a row, a Writer's Workbench program flags the double entry even if it is split across two lines. This is a

*UNIX is a trademark of Bell Laboratories.

TABLE 1. STRAIGHTFORWARD FEATURES

Feature	Writer's Workbench	CRES	WRITEAIDS
1. Spelling Check	Internally Developed	Limited to Master Word List	Commercial software
2. Punctuation Check	X	Not planned	Not planned
3. Double Word Check	X	Not planned	Not planned
4. Sexual Bias	About 170 words and phrases mechanized	Planned for Mil-Spec compliance	Planned for Mil-Spec compliance
5. Parts of Speech Analysis	% of various type of words used	Planned	Not planned
6. Sentence Type	% of various types of sentences used	Planned	Not planned

difficult error for humans to detect.²

Double word checks are not planned for CRES or WRITEAIDS.

2.1.4 Sexist Language

The Writer's Workbench has a program that identifies terms that might suggest sexual bias in writing.² The program contains a dictionary of about 170 words and phrases that may be sexist.

Future versions of CRES and WRITEAIDS will have sexual bias programs that comply with military specifications for technical manuals.

2.1.5 Parts of Speech Analysis

A parts of speech analysis is performed by a set of Writer's Workbench programs.⁵ These programs use a small dictionary, suffix rules, and experimentally derived word order rules to assign word classes to all words in a text. The analysis is about 95 percent

accurate. One purpose of these programs is to provide statistics on the percentage and raw count of various word types used in a text. These programs are also used with several other Writer's Workbench programs to be described later.

Parts of speech identifiers are coded with the entries in the CRES Master Word List. Plans for using this data are not yet complete.⁶

Parts of speech data is built into an experimental WRITEAIDS program for computer-aided translation. There are no plans to use such data for other purposes.

2.1.6 Sentence Type Analysis

A Writer's Workbench program performs a sentence-type analysis. The program determines the raw count and percentage of simple, complex, compound, and compound-complex sentences in a document.⁵

A sentence type analysis feature is planned for CRES⁶ but not for WRITEAIDS.

2.2 SUGGESTIONS FROM WRITING EXPERTS

Table 2 summarizes automated features that are based on suggestions from writing experts concerning grammar or clear writing principles.

2.2.1 Split Infinitives Check

The use of split infinitives is often considered to be a grammatical error. Using the parts of speech analysis program described in Paragraph 2.1.5, the Writer's Workbench will find infinitives that are

split by adverbs.² Users can obtain printed grammatical information about split infinitives from another program.

A split infinitive check is not planned for CRES⁶ or WRITEAIDS.

2.2.2 "To Be" Check

Some writers, including Lanham, suggest that document revisors should circle all forms of the verb "to be" and try to replace them.⁷ Following this advice could help cut down on the use of passive voice,

TABLE 2. SUGGESTIONS FROM WRITING EXPERTS

Feature	Writer's Workbench	CRES	WRITEAIDS
1. Split Infinitive Check	Flags infinitives split by adverbs	Not planned	Not planned
2. "To Be" Check	Flags all forms of "to be"	Passives only	Not planned
3. Preposition Count	Included in parts of speech analysis	Identifies excessive use of prepositions in a sentence	Not planned
4. Wordiness/Phrase Substitution	About 400 phrases mechanized	About 250 phrases mechanized	About 350 phrases compiled; software under development
5. Sentence Opener Check	X	X	Expletive check Planned
6. Clause Length Check	Not Planned	X	Not planned
7. Sentence Type Variability Check	X	Not planned	Not planned
8. Paragraph Length Check	Not planned	X	Planned
9. Organization Check	X	Not planned	Planned

nominalizations (nouns made from verbs), and expletives (words that add no information).

A Writer's Workbench program underlines and capitalizes all forms of "to be". The text printout is formatted normally so potential problems are highlighted in context. Bell reports that this is a useful way to look at the first draft of a paper.⁴

The CRES system is not specifically programmed to find all forms of "to be", but it will locate full passive constructions. A "to be" check is not planned for WRITEAIDS.

2.2.3 Preposition Count

The excessive use of prepositions may indicate wordiness. One output of the Writer's Workbench parts of speech analysis is a preposition count. This data is probably most useful when included in a style comparison (refer to Paragraph 2.3.11).

CRES flags sentences containing more than a predetermined number or percentage of prepositions.⁸ A preposition count is not planned for WRITEAIDS.

2.2.4 Wordiness/Phrase Substitution

Writing experts have identified several phrases as wordy, awkward, repetitive, or frequently misused. Many such phrases can be replaced with a single word or a shorter phrase.

The Writer's Workbench has a multi-purpose dictionary containing about 400 wordy or frequently misused phrases.⁵ String matching programs flag the faulty phrases and suggest one or more substitutes. The writer or editor must decide if a suggested substitute is appropriate in the context in which it would be used. Users can tailor the

dictionary to their own needs by adding or deleting terms.

CRES also has a multi-purpose dictionary containing about 250 phrases with as many as two suggested substitutes per phrase.^{6,8} As with the Writer's Workbench, CRES users must decide whether suggested substitutes are appropriate. CRES users can tailor the phrase dictionary to their own needs.

A dictionary containing about 350 phrases with one or more substitutes per phrase has been compiled for WRITEAIDS.⁹ The purpose and use of this dictionary are virtually identical to the Writer's Workbench and CRES. Software to implement this feature is under development.

2.2.5 Sentence Opener Check

Writing experts have made two kinds of suggestions about sentence openers. The first is that variety in sentence openers is desirable. The second is that some types of sentence openers are wordy or should be avoided in certain instances.¹⁰

Using the parts of speech analysis programs, the Writer's Workbench provides statistics on the percentage and raw count of the types of sentence openers used in a text.⁵ With experience and the aid of an interpretive program, these statistics can be used to achieve a desirable variety in sentence openers and to avoid the excessive use of undesirable sentence openers such as expletives.

As part of its phrase dictionary software, CRES identifies expletives. Another CRES program advises users to avoid starting sentences with certain subordinate clauses.⁸

The only sentence opener check planned for WRITEAIDS is an expletive check.

2.2.6 Clause Length/Awkwardness Check

Experts, including Klare,¹¹ advise that clause length is important apart from sentence length. They suggest reducing clause length where possible.

A clause length check is not planned for the Writer's Workbench. CRES flags clauses with more than 10 words and flags awkward dependent clauses.⁸ A clause length check could be programmed for WRITEAIDS but is not now planned.

2.2.7 Sentence Type Variability Check

Many experts agree that sentence variety adds to reading pleasure and can help hold a reader's attention.^{11,12} Some writers do not realize that their sentences lack variety. Faced with stringent format and style requirements, they may find it difficult to consider sentence variability.

Statistics on sentence variety are provided by the Writer's Workbench.⁵ Interpretive programs provide writers with both positive and negative variability characteristics of their text. The program provides numerous suggestions for increasing sentence variety.²

Sentence type variability checks are not planned for CRES⁶ or WRITEAIDS.

2.2.8 Paragraph Length Check

Long paragraphs can be hard to understand. This is perhaps because the content drifts too far from the main topic or because the writer is poorly organized. A long paragraph check can draw attention to paragraphs that may need to be rewritten.

Paragraph length checks are not planned for the Writer's Workbench.

The CRES system flags paragraphs that exceed 10 sentences.⁸ A paragraph length check based on a word count is planned for WRITEAIDS.

2.2.9 Organization Check

The organization of a document provides readers with useful clues that aid understanding. Until the 1970's, there was virtually no useful research on organization, so writers relied on the advice of experts. Those experts often recommended that the first sentence of a paragraph should be a topic sentence or should serve as a transition from the previous paragraph to a new topic. While this is sound advice, organization is related to content in other complex ways. The recent research on organization is still quite theoretical¹¹ and can not yet be automated. Hopefully, future organization suggestions will move into the category of suggestions based on research.

The Writer's Workbench has a useful program to help writers evaluate organization. The program prints paragraph headings and preserves paragraph boundaries, but it prints out only the first and last sentence of each paragraph. This printout can provide an abstract of a paper for writers who use traditional formats; it provides a structural outline for writers with a more casual style.²

There are no plans to include an organizational check in the CRES system.

In 1983, research on the practical application of comprehension, coherence, and inference loading theories will begin at Westinghouse. The long range goal is to implement research-based organizational checks in the WRITEAIDS system.

2.3 SUGGESTIONS BASED ON RESEARCH

In this category, the similarities of the three systems are pronounced. The differences between the systems tend to reflect either the organizational goals of Bell Laboratories, TAEG, or Westinghouse or the areas of most interest to the various researchers. There is no doubt that the three organizations are working with the same basic ideas and data. Differences in available hardware and software have led to some of the different approaches for applying the suggestions based on research. Table 3 summarizes automatic features based on research in readability, comprehensibility, psychology, linguistics, and psycholinguistics.

2.3.1 Readability Analysis

The term readability analysis, as applied here, refers to predicting the style difficulty of text by readability formula. Readability formulas can be misused because they are not intended to measure content, organization, format, legibility, or other factors considered important to comprehension. Laymen often drop the word "readability" from formula score descriptions and thus tend to equate the scores with age groups or conceptual difficulty, neither of which is correct.

When the data from a readability analysis is properly applied, it can provide valuable insight into potential problems with word use, sentence use, or both. When a thoughtful sampling procedure is used, a readability analysis will also highlight the particular portions of a document that are most likely to need revision. While perhaps hundreds of readability formulas exist, all three organizations agree that the Flesch-Kincaid formula seems to be the most appropriate formula for technical material.^{5,3,14}

The Writer's Workbench calculates four readability indices: The Flesch Reading Ease Score¹⁵, The Flesch-Kincaid Formula¹⁶, The Automated Readability Index¹⁷, and The Coleman-Liau Formula¹⁸. One program prints the four formula scores and provides statistics on the average word and sentence length. Another program interprets the Flesch-Kincaid score for users.²

CRES calculates the Flesch Reading Ease Score and the Flesch-Kincaid Formula. The CRES readability analysis provides statistics on the average number of words per sentence and the average number of syllables per word. It also lists the words in a text which do not appear on the Master Word List (refer to Paragraph 2.3.4).

WRITEAIDS also calculates the Flesch Reading Ease Score and the Flesch-Kincaid Formula. The WRITEAIDS readability analysis provides raw data and the same statistical data as CRES. WRITEAIDS also lists words of three or more syllables in the order that the words appeared in the analyzed text.

2.3.2 Test Question Analysis

Research has identified numerous principles for constructing good true-false and multiple-choice test questions. For a summary of that research, see Kincaid, Braby, and Wulfeck.¹⁹ TAEG has a specific need for test question analysis so CRES is the only one of the three systems that automates this feature.

The CRES test question analysis includes all of the regular features of CRES plus several features specified by the Navy's Instructional Quality Inventory²⁰ (IQI). IQI features for multiple-choice questions include flagging long sentences, inappropriate answers, the fact that the longest answer is correct, and repetitive words and

TABLE 3. SUGGESTIONS BASED ON RESEARCH

Feature	Writer's Workbench	CRES	WRITEAIDS
1. Readability Analysis	Flesch, Flesch-Kin. ARI, Col-Liau	Flesch, Flesch-Kin.	Flesch, Flesch-Kin.
2. Test Q Analysis	Not Planned	Readability, item construction	Not planned
3. Readability Sum.	X	Meets Mil-spec	Meets Mil-spec
4. Vocabulary Check	Not planned	4300 root words, plus supplements	Two vocabs. planned
5. Word Substitution	100 root words, less awkward substitutes	250 root words, substitutes based on length	1000 root words, substitutes based on length, fam., freq.
6. Abstract Word Check	Nouns, verbs, adjs.	Planned	Nouns-verbs, adjs. soon
7. Nominalization Check	X	X	Exception List
8. Long Sen. Ck.	X	X	X
9. Neg. Statement Check	Not planned	Double negs. text - all negs. tests	Direct negs.
10. Passive Check	X	Full passives	Under devel.
11. Style Comparator	X	Not planned	Future

phrases. IQI features for true-false questions include flagging long questions, negative wording, and complex questions.¹⁹

2.3.3 Readability Summary of Several Passages

A readability summary of the analysis of several passages from the same document helps writers identify specific types of writing or specific parts of a document that are most likely to need revision.

Interpretive programs in the Writer's Workbench provide both tabular and written summaries of the scores for individual passages. Users can compare the summaries for the individual passages. These programs also provide written information on the background that readers might need to understand the document.²

Both CRES⁶ and WRITEAIDS provide readability summaries that permit a side-by-side comparison of the

passages within a document. These summaries include the raw data, index variables, and readability scores for individual passages. Both systems also provide an overall readability analysis of an entire document that complies with military specification requirements.

2.3.4 Controlled Vocabulary Check

Studies show that when the words used in written materials are familiar to intended readers the materials are easier to read and understand. Several researchers have prepared "controlled vocabularies" which permit the use of only those terms that are either known to be familiar or will be taught to intended readers. Many "controlled vocabularies", ranging in length from about 1000 to 4000 root words, have been prepared for technical materials. Two types of "controlled vocabularies" have emerged. The first type is based on word frequency analysis and permits multiple meanings of the same word. The second type is based on the one word-one meaning concept and permits only one shade of meaning for a word.

There are no plans to include a controlled vocabulary in the Writer's Workbench.¹³

CRES has a controlled vocabulary called the Master Word List. This list is based on (1) word frequency analysis of Navy training materials, and (2) word lists prepared by several military and nonmilitary organizations.³ The Master Word List consists of about 4300 root words plus all inflected forms of the root words. The Master Word List can be tailored to specific audiences by adding supplementary technical terms that are peculiar to special disciplines. CRES flags all words in a text that do not appear on the Master Word List. A writer or editor replaces flagged words with common words where possible.

Two "controlled vocabularies" are planned for WRITEAIDS. The plans are to add the CRES word lists and an internally developed one word-one meaning logistics glossary.

2.3.5 Word Substitution

Readability research has shown that a reader's speed, recall, and comprehension can often be improved if the difficult content words of a passage are changed in certain ways. Words that are shorter, more frequently used, or more familiar tend to improve human performance in reading. See Klare¹¹ for an extensive discussion of the literature. Word substitution dictionaries are designed to suggest simpler alternatives for the difficult nontechnical words that appear in a passage. Before using a suggested substitute, the writer or editor must make sure that the meaning of the text will not be changed.

The same Writer's Workbench dictionary that lists awkward or misused phrases also contains about 100 words that are potentially awkward. A string matching program flags the difficult words when they occur in a passage. Users can request a printout that suggests one or more shorter substitutes for the flagged words. Users can also tailor the dictionary to their own needs by adding or deleting terms.⁴

The CRES word substitution dictionary contains about 250 potentially difficult root words.⁶ CRES automatically flags each word in a passage that matches a word in the dictionary and prints out up to two suggested substitutes.³ Terms can be added to or deleted from the dictionary.⁸

The WRITEAIDS substitution dictionary contains about 1000 potentially difficult words. WRITEAIDS automatically flags each word in a

passage that matches a word in the dictionary and prints out all available suggested substitutes. The dictionary was specially constructed to suggest changes based on readability research. With a few minor exceptions, each suggested substitute is a shorter word, a more familiar word²¹, and a more frequently used word.²² When all appropriate word changes are inserted, the readability grade level of a document is automatically lowered; the probability is that the comprehensibility of the document is also improved.

2.3.6 Abstract Word Check

Concrete words, which easily arouse an image in one's mind, contribute more to readable writing than abstract words. Studies have shown that concreteness can improve human performance scores in recognition and recall memory, readability judgements, sentence usage, information gain, and recognition time. See Klare¹¹ for a summary of the literature.

The conceptual abstractness of text is checked by a Writer's Workbench program. The program calculates the percentage of words in a text that appear on a list of 314 abstract nouns, verbs and adjectives compiled through psychological research. When the percentage of abstract words exceeds 2.3 percent, the program suggests that concrete examples should be used to clarify the text.⁴

An abstract word check is planned for CRES.

WRITEAIDS contains a list of about 250 abstract words. A string matching program flags each word in the text that occurs on the abstract word list. A user's manual suggests that concrete examples should be used to clarify the text.²³ The abstract word list now contains only

nouns; verbs and adjectives will soon be added.

2.3.7 Nominalization Check

Excessive use of nominalizations (nouns made from verbs) tends to make text less readable. Studies have shown that replacing nominalizations with the active verb form can improve human performance scores in comprehension, recall, and information gain. See Klare¹¹ for a summary of the literature.

The Writer's Workbench program for parts of speech analysis identifies nominalizations. Another program calculates the raw count and percentage of nominalizations in the text. An interpretive program advises writers to avoid excessive nominalizations by transforming nominalized sentences to the active verb form.²

A nominalization check is available for CRES.⁶

The WRITEAIDS nominalization check is based on the same string matching techniques used in the Writer's Workbench. A special exception dictionary is used to avoid flagging words which are actually verbs or are both nouns and verbs. The program flags each nominalization that occurs in a passage. A user's manual advises writers to transform nominalizations to the active verb form where possible.²³

2.3.8 Long Sentence Check

Shortened sentences contribute to more readable writing. Research shows that certain sentence changes improve human performance in reading comprehension and speed. See Klare¹¹ for a discussion of the literature.

A Writer's Workbench program determines the percentage of short and long sentences and the length

and location of the shortest and longest sentences in a passage.⁵ Another program allows users to locate all sentences that exceed a specified number of words or contain less than a specified number of words.

The long sentence checks in CRES and WRITEAIDS are virtually identical. Both systems automatically flag the end of each sentence that exceeds a specified number of words.

2.3.9 Negative Statement Check

Negative statements sometimes require more time or cause more errors in verifying truth or falsity than affirmative statements.¹¹ Many negative statements can be transformed into affirmative statements.

A negative statement check is not planned for the Writer's Workbench.

CRES offers two types of negative statement checks. For normal text, CRES flags each sentence containing a double negative. When used to check true-false or multiple-choice test questions, CRES flags any negative construction.

WRITEAIDS flags all occurrences of the direct negatives no, not, or nor. Numerous additional forms of negative constructions have been identified but have not yet been programmed. A user's manual asks writers to transform negative statements to positive statements where possible.²³

2.3.10 Passive Voice Check

The excessive use of passive sentences is one of the biggest problems in technical writing. Sentences written in passive voice tend to be longer, less direct, and less vigorous than sentences written in active voice.²⁴ Changing sentences from passive voice to

active voice can improve a reader's verification and recall.¹¹ However, passives are often needed to add variety, emphasize the object of the sentence, or avoid naming an unknown or unimportant actor.

The Writer's Workbench calculates the percentage of passives used and will locate passive constructions on request. An interpretive program advises users on how to revise sentences to avoid the excessive use of passive voice.^{2,5}

The CRES system flags each passive sentence that can be transformed to active voice.⁸ CRES, therefore, avoids flagging truncated passives which are often impossible to transform. A passive voice check, similar to that in the CRES system, is planned for WRITEAIDS.

2.3.11 Style Comparator

A style comparator not only provides style statistics but it also interprets the statistics. This feature compares the style of a new text against a set of style standards and describes the differences.

Bell Laboratories has prepared several sets of style standards for various types of documents. The style standards reflect the writing from a collection of documents of each type, which were judged to be good.² A Writer's Workbench user selects which set of built-in standards should be used to interpret a new document. Users can also tailor standards for a particular audience by creating new standards from a set of documents. The Writer's Workbench style comparator identifies those style values of a new document which are outside the range of the selected standard. A two-to-three page output in written English explains why deviations from the standards may make the new document hard to

understand. The same output explains how to rewrite the text to remove problems. Users receive positive feedback when style values meet the built-in standards.^{2,4}

A style comparator is not planned for CRES.⁶ Funding for development of a WRITEAIDS style comparator has been requested.

3. FEEDBACK PROVIDED BY THE SYSTEMS

The word feedback, as used in this section, refers to the information supplied by the systems to the users. All of the systems allow users to select various program options from a menu; in some instances there are both short and long versions of the same option. The systems provide feedback on either a hard copy printout or a CRT display. This paper will only describe feedback from hard copy printouts.

To obtain data for this section, ten samples of text were analyzed by each of the three systems. Space does not permit showing all of the data (over 150 pages). However, the Appendix contains all of the printouts provided by each system for one of the ten samples. These printouts provide representative examples of many of the text analysis features described in Section 2. Differences between the comparable features of the three systems reflect differences in the algorithms, word lists, and program formats.

3.1 WRITER'S WORKBENCH

The feedback provided by the Writer's Workbench can be grouped into three categories. Those categories are described as statistical analysis, objective analysis, and interpretive analysis. The Appendix shows the Writer's Workbench printouts in the order in which they are discussed.

3.1.1 Statistical Analysis

The Writer's Workbench provides statistical feedback on a wide range of style variables. For a given document, the variables reported include readability indexes, information on average word and sentence lengths, the distribution of sentence lengths and types, the usage of various parts of speech, and the distribution of sentence openers. The data includes both raw counts and percentages. Standing alone, the output of the program that compiles statistical data is most useful for research on the style of documents. However, the output of the program is also used as the input for the interpretive analysis described in Paragraph 3.1.3.

3.1.2 Objective Analysis

The Writer's Workbench performs several tedious and unrewarding tasks that were traditionally done by writers and editors. Proofreading programs check for problems which are objective in nature such as spelling, punctuation, grammar, and sexual bias. An organization program provides a look at the overall structure of a document.

Feedback from the proofreading analysis is provided by English language printouts. The printouts first describe the type of check performed (spelling, punctuation, double word entry, word choice, split infinitives, or sexist language). Next, the printouts identify each possible error found; if no error is found, a brief statement to that effect is printed out. Finally, for each possible error detected, the printout lists suggested corrections or revisions. The decision to accept or reject the suggestions is left to the user.

The organization program preserves the headings and paragraph

boundaries of the text, but it prints out only the first and last sentence of each paragraph. The actual evaluation of the text structure is left to the user.

3.1.3 Interpretive Analysis

The output from programs for statistical style analysis is used as the input for interpretive programs available on the Writer's Workbench. The interpretive programs provide users with positive and negative feedback in one of two forms. The longer form provides a written report that discusses in prose how the input text compares with a set of built-in standards. The short form compares the input text with the same standards but prints out only brief comments.

For short documents, the long form starts with a one sentence warning that the interpretive analysis may be misleading. A one sentence note then informs the user of the standards being used to judge the input text. Next, the printout provides a written comparison of the style variables of the input text with the built-in standards. If a particular style variable meets the standard, a short statement to that effect is printed out. If a variable (readability for example) exceeds the standard, the user is provided with a longer report containing remedial information. That information describes the problems that readers may encounter because the standard was exceeded, some reasons why the standard was exceeded, and practical suggestions for improving the text.

The short form of the interpretive analysis also starts with a warning (for short documents) that the analysis may be misleading. It then lists the style variables analyzed and simply states whether the score for each variable is good or high.

If the score is high, the printout suggests what a good score would be and perhaps a brief suggestion for improvement.

3.2 CRES

CRES provides two types of feedback. The first is a statistical analysis and the second is an "in context" analysis. The Appendix shows both types of feedback.

3.2.1 Statistical Analysis

CRES provides statistical feedback on readability and word usage for a particular document. A printout shows the raw count of sentences, words, and syllables, information on average word and sentence lengths, readability indexes, and an alphabetical listing of uncommon words. This type of feedback is useful for evaluating compliance with requirement specifications and highlighting potential areas of style difficulty.

3.2.2 "In Context" Analysis

CRES provides feedback on style analysis by directly flagging potential problems in the context in which they actually occur. Potential problems with particular words or phrases are flagged by bracketing the actual word or phrase with a set of symbols. Uncommon words are bracketed by the symbols "[]". Words or phrases for which there are suggested substitutes are bracketed by the symbols "<>"; suggested substitutes are inserted immediately after flagged items using the same set of symbols. The number of words in a long sentence is printed out within the "<>" symbols after the sentence identifying punctuation. All comments about style are made within the "<>" symbols as potential problems occur anywhere in the text. A CRES user's manual explains the meaning of the symbols and provides practical advice.

improving text in response to the flagged problems.

3.3 WRITEAIDS

WRITEAIDS also provides two types of feedback. Like CRES, the first is a statistical analysis and the second is an "in context" analysis. The Appendix shows both types of feedback.

3.3.1 Statistical Analysis

WRITEAIDS provides statistical feedback on the readability of a document. A printout shows the raw count of sentences, words and syllables, information on average word and sentence lengths, and readability indices. The same printout also lists the long words in the order in which they occur in the document. Readability statistics and long word lists are useful for evaluating compliance with requirement specifications and highlighting potential areas of style difficulty.

3.3.2 "In Context" Analysis

Like CRES, WRITEAIDS provides feedback on style analysis by identifying potential problems in context. Unlike CRES, comments and suggested changes are made below each line of text instead of being interspersed in the text. Two different types of printouts are used to provide "in context" feedback.

One printout shows suggestions for word substitution and identifies long sentences. This printout shows each line of text exactly as it appears in the document being analyzed. Below each line of text is a list of the words for which there are suggested substitutes; all available suggested substitutes are printed out immediately after each listed word. If a line of text includes the end of a long sentence, a "long sentence" note is printed out below the line. Users can mark

accepted word and sentence changes directly on this printout.

The program that produces the printout described in the preceding paragraph also creates a special file (stripped of punctuation) used in other style analysis checks. This file is printed out double spaced. The blank lines below each line of text contain flags that identify potential style problems. For example, the flag NOM appears directly beneath each nominalization. Multiple flags can be printed out for the same word. A user's manual explains the flags and makes practical suggestions for improving the text.

4. USER ACCEPTANCE

Because large scale tests of user acceptance began only recently, there is very little hard data to report on this topic. However, users in general report that the systems are useful, serve their intended purpose, improve the quality of final documents, and may improve the skills of writers and editors. Users also seem to appreciate the impersonal nature of the suggestions made by the systems.

4.1 WRITER'S WORKBENCH

One survey of Writer's Workbench users found that most think the feedback from the system is clear, and the system is likely to find things the user would miss. A significant number of users reported that the system improves their writing skills. Other trials have shown that under time pressure, technical writers detected significantly more planted errors with the system than without it.²

4.2 CRES

CRES has been used by two large defense contractors to edit technical manuals. Users report that

the use of CRES results in improved documentation with about a 50 percent savings in editing time.⁶

4.3 WRITEAIDS

Experiments using WRITEAIDS to edit technical manuals began in September 1982, so only the general comments made in Paragraph 4 can be made.

5. FUTURE APPLICATIONS

Computer-aided editing systems already provide a wealth of information and numerous ways to apply that information. Future applications or adaptations of the systems seem to be limited only by imagination. Some future uses seem obvious once they are stated; others require insight into the problems faced by readers, writers, and editors. The closing paragraphs of this paper will explore, without elaboration, a wide range of possible applications.

5.1 ADDITIONAL FEATURES

Research on reading and writing has provided some suggestions for improving text that have not yet been automated. Some of these suggestions that could be additional features include:

- (1) a check for references back to earlier parts of a text (anaphora),
- (2) a check for references forward to later parts of a text (inferences),
- (3) a word association check,
- (4) a word depth check, and
- (5) a self-embedded statement check.

Future research on reading and writing may provide additional text analysis features. Some possible features are:

- (1) an index of conceptual difficulty,
- (2) a check on the coherence of a document, and

- (3) a measure or check of organization.

Additional objective features that could save production time are:

- (1) automatic generation of an index, and
- (2) automatic generation of a table of contents (this feature is available at Bell Laboratories).

5.2 ADAPTATION TO OTHER USES

With little or no further development, computer-aided editing can be adapted to other uses. Possible adaptations include:

- (1) editing the data base used in computer-aided authoring systems,
- (2) using the text analysis features to help writing students analyze their own texts (experiments using the Writer's Workbench are being performed),
- (3) using style comparators to help students compare their writing with that of famous writers or with standards,
- (4) using style comparators to smooth out differences in style between multiple contributors to the same document, and
- (5) using text analysis to help instructors diagnose problems in students' writing.

With further development, software techniques for editing could be combined with software for other applications to help solve problems in other areas. For example:

- (1) editing software, could be combined with software that divides sentences at pausal boundaries to prepare more understandable CRT messages (this would be especially useful for preparing emergency operating procedures), and
- (2) software for parts of speech analysis, word substitution, and phrase substitution could be

combined with software for syntax conversion to prepare fairly high quality translations from English to other languages.

ACKNOWLEDGEMENTS

Many people helped me to prepare this paper. Larry Frase and Nina Macdonald of Bell Laboratories provided several articles about the Writer's Workbench, answered numerous questions, and reviewed the paper for technical content. Mary Fox provided the data from the samples of text she ran on the Writer's Workbench programs. Peter Kincaid of TAEG provided documentation on CRES, answered numerous questions, and reviewed the paper for technical content. Ray Cox provided the data from the samples he ran on CRES. Edward Pierce of Westinghouse ran all of the text samples through WRITEAIDS. Shirley Bloom did all of the typing; Charles Lighthiser and Frank Wojcicki did the editing.

A special acknowledgement is due to the programmers and software designers who have made computer-aided editing a reality. Major contributors at Bell Laboratories

include Lorinda Cherry, Stacey Keenan, and Nina Macdonald. Early contributors at TAEG were Nora Gregory, Carolyn Trotta, and Charles Guitard. Ray Cox has been the principal programmer for the last two years. Mr. Edward Pierce prepared all of the computer-aided editing software at Westinghouse.

BIOGRAPHY

J. Douglas Kniffin is a Fellow Engineer with the Technical Data and Training Systems Department, Integrated Logistics Support Divisions, Westinghouse Electric Corporation. He is currently Principal Investigator for IR&D in the area of Technical Manual Technology Research. He participated in the Westinghouse-Johns Hopkins Scholarship Award Program and received his BSEE from Johns Hopkins University. He also holds a Doctorate in Law from the University of Maryland. Recently he has been Principal Investigator for several R&D programs in readability/comprehensibility for both Westinghouse and the military services. Doug is currently President of the Metro-Baltimore Chapter of the Society for Technical Communication.

REFERENCES

1. M. D. McIlroy, "Development of a Spelling List," IEEE Transactions on Communications, Special Issue on Communications in the Automated Office, 30, No. 1 (January 1982), pp 91-99.
2. N. H. Macdonald, L. T. Frase, P. S. Gingrich, and S. A. Keenan, "The Writer's Workbench: Computer Aids for Text Analysis," IEEE Transactions on Communications, Special Issue on Communications in the Automated Office, 30, No. 1 (January 1982), pp 105-110.
3. J. P. Kincaid, J. A. Aagard, and J. W. O'Hara, "Development and Test of a Computer Readability Editing System (CRES)," TAEF Report No. 83, Orlando, Florida (March 1980).
4. N. H. Macdonald, "The Writer's Workbench: Rationale and Design," Bell System Technical Journal, Special Issue on Human Factors and Behavioral Science (In Press).
5. L. L. Cherry, "Writing Tools," IEEE Transactions on Communications, Special Issue on Communications in the Automated Office, 30, No. 1 (January 1982), pp 100-105.
6. J. P. Kincaid, TAEF, (personal communication).
7. R. A. Lanham, Revising Prose, New York: Charles Scribners' Sons (1979).
8. R. Cox, "Computer Readability Editing System Users Manual", TAEF, Orlando, Florida (June 1982).
9. J. D. Kniffin, "Phrase Substitution Dictionary" Westinghouse Electric Corporation, (Unpublished), (1982).
10. J. A. Brogan, Clear Technical Writing, New York: McGraw-Hill Inc., (1973), pp 147-148.
11. G. R. Klare, A Manual for Readable Writing, Glen Burnie, Maryland: REM Company, (1980).
12. D. W. Ewing, Writing for Results, New York: John Wiley & Sons, Inc., (1974).
13. L. T. Frase, Bell Laboratories, (personal communication).
14. J. D. Kniffin, "The New Readability Requirements for Military Technical Manuals," Technical Communication, (Third Quarter 1979), pp 16-19.
15. R. Flesch, "A New Readability Yardstick," Journal of Applied Psychology, Vol. 32, (1948), pp 221-233.
16. J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chisom, "Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel," Navy Training Command Research Branch Report 8-75, (1975).

17. E. A. Smith and J. P. Kincaid, "Derivation and Validation of the Automated Readability Index for Use with Technical Manuals," Human Factors, Vol. 12, (1970), pp 457-464.
18. M. Coleman and T. L. Liau, "A Computer Readability Formula Designed for Machine Scoring," Journal of Applied Psychology, Vol. 60, (1975), pp 283-283.
19. J. P. Kincaid, R. Braby, and W. H. Wulfeck, "Computer Aids for Editing Test Questions," Educational Technology, (in press), (1982).
20. P. S. Fredericks, "The Instructional Quality Inventory III Training Workbook", NPRDC Special Report 80-25, San Diego, California, (1980).
21. E. Dale and J. O'Rourke, The Living Word Vocabulary, Elgin, Illinois: DOME Inc., (1976).
22. H. Kucera and W. N. Francis, Computational Analysis of Present-Day American English, Providence, Rhode Island: Brown University Press, (1967).
23. E. J. Pierce and J. D. Kniffin, "WRITEAIDS User's Manual", (unpublished).
24. R. Gunning, The Technique of Clear Writing, New York: McGraw-Hill Book Company, (1968).

APPENDIX 1 TO CHAPTER 12: WRITERS WORKBENCH

Aug 24 13:05 1982 THIS IS THE STYLE TABLE FOR kniffin06 Page 1

readability grades:

(Kincaid) 16.2 (auto) 16.5 (Coleman-Liau) 15.2 (Flesch) 17.0 (22.2)

sentence info:

no. sent 19 no. wds 463

av sent leng 24.4 av word leng 5.48

no. questions 0 no. imperatives 0

no. content wds 262 55.6% av leng 7.56

short sent (<19) 16% (3) long sent (>34) 11% (2)

longest sent 37 wds at sent 2; shortest sent 6 wds at sent 6

sentence types:

simple 26% (5) complex 47% (9)

compound 11% (2) compound-complex 16% (3)

word usage:

verb types as % of total verbs

tobe 30% (16) aux 21% (11) inf 26% (14)

passives as % of non-inf verbs 13% (5)

types as % of total

prep 12.3% (57) conj 1.9% (9) adv 3.7% (17)

noun 26.3% (122) adj 16.8% (78) pron 3.9% (18)

nominalizations 4 % (20)

sentence beginnings:

subject opener: noun (0) pron (1) pos (0) adj (4) art (4) tot 47%

prep 16% (3) adv 26% (5)

verb 0% (0) sub_conj 11% (2) conj 0% (0)

expletives 0% (0)

Aug 24 13:01 1982 PROOF OUTPUT FOR kniffin06 Page 1

***** SPELLING *****

Possible spelling errors in kniffin06 are:

anaphora

If any of these words are spelled correctly, later type
spelladd word1 word2 ... wordn
to have them added to your spelldict file.

***** PUNCTUATION *****

The punctuation in kniffin06 is first described.

4 double quotes and 0 single quotes
1 apostrophes
0 left parentheses and 0 right ones

The program next prints any sentence that it thinks is
incorrectly punctuated and follows it by its correction.

No errors found in kniffin06

***** DOUBLE WORDS *****

For file kniffin06:

No double words found.

***** WORD CHOICE *****

Sentences with possibly wordy or misused phrases are listed next,
followed by suggested revisions.

For file kniffin06

beginning line 1 kniffin06
Because our *[endeavor]*s to ameliorate readability and text
processing are familiar at numerous echelons of the military,
the requirement for an alteration in our standard demonstration
therefore became obvious.

beginning line 9 kniffin06
Accordingly, you are observing an amalgamation of the products
of our ongoing and future research with our objective being to
exhibit a consolidated demonstration of our efforts at producing
beneficial *[modification]*s in both text and text processing.

beginning line 14 kniffin06

Aug 24 13:01 1982 PROOF OUTPUT FOR kniffin06 Page 2

We *[anticipate]* that *[subsequent]* to the termination of this demonstration, the human propensity to forget all the technicalities will take effect; however, by retaining the sample transcriptions provided to you, you can *[facilitat]*e your memory.

beginning line 21 kniffin06

Today we are exhibiting *[one of the]* primary features of our new computer-aided editing system; this feature has been *[designat]*ed "the substitution dictionary.

beginning line 31 kniffin06

One list *[compris]*es *[approximately]* 900 difficult words that frequently appear in technical manuals while the second list *[compris]*es *[approximately]* 1800 suggested substitutes for the words in the original list.

beginning line 35 kniffin06

Each substitute was selected to match the shade of meaning of the difficult word that it might replace, and additionally each substitute is a shorter, more familiar, and more frequently *[utiliz]*ed word.

beginning line 42 kniffin06

The text processor is then *[utiliz]*ed to compare each word in the text with the list of difficult words in the substitution dictionary.

beginning line 49 kniffin06

The writer must then determine *[which]* word *[modification]*s can be made without producing an alteration in the meaning of the text.

beginning line 52 kniffin06

A readability analysis of the edited text can then be *[accomplished]* to determine if the text is at an appropriate level of difficulty for the intended audience.

beginning line 56 kniffin06

[In the near future], we hope to add numerous innovations to the computer-aided editing system.

beginning line 59 kniffin06

[At this time,] we believe that *[desirable benefits]* will result *[by means]* of taking advantage of characteristics that are peculiar to certain structures.

beginning line 63 kniffin06

Other benefits could *[follow after]* we *[conduct an investigation]* of the *[manner]* in *[which]* writers make references to other parts of their own writing.

beginning line 66 kniffin06

For example, we believe that the identification of passive voice, anaphora, and inference loading may be *[practicable]*.

beginning line 70 kniffin06

Aug 24 13:01 1982 PROOF OUTPUT FOR kniffin06 Page 3

Furthermore, we are planning to *[implement]* the U.

beginning line 71 kniffin06

Navy's 4000 word controlled vocabulary and to *[fabricate]* our own "Controlled Logistics Vocabulary.

beginning line 72 kniffin06

" Finally, if we *[couple together]* the research on word and phrase substitution, parts of speech analysis, and syntax roles, we will have the *[necessary requisite]*s for *[subsequent]* translation from English to foreign languages.

file kniffin06: number of lines 77 number of phrases found 30

----- Table of Substitutions -----

PHRASE

SUBSTITUTION

accomplished: use "done" for " accomplished"
 anticipate: use "expect" for " anticipate"
 approximately: use "about, nearly, almost" for " approximately"
 at this time: use "now" for " at this time"
 by means: use "with, by, through" for " by means of"
 compris: use "includes" for " comprise"
 conduct an investigation: use "investigate" for " conduct an investigation of"
 couple together: use "couple" for " couple together"
 designat: use "appoint, name" for " designate"
 desirable benefits: use "benefits" for " desirable benefits"
 endeavor: use "try" for " endeavor"
 fabricate: use "hand made, manufactured, fabricated" for " man made"
 fabricate: use "make, build" for " fabricate"
 facilitat: use "ease, simplify, help, improve" for " facilitate"
 follow after: use "follow" for " follow after"
 implement: use "execute, fulfill, achieve" for " implement"
 in the near future: use "soon" for " in the near future"
 manner: use "like, as" for " after the manner of"
 manner: use "precisely, etc." for " in a precise manner"
 manner: use "similarly" for " in a similar manner"
 manner: use "way" for " manner"
 modification: use "change" for " modification"
 necessary requisite: use "requisite" for " necessary requisite"
 one of the: use "one, a" for " one of the"
 practicable: use "possible" for " practicable"
 subsequent: use "later, after" for " subsequent to"
 subsequent: use "next, later, further" for " subsequent"
 utiliz: use "use" for " utilize"
 which: use ""that" when clause is restrictive" for " which"
 which: use "of which" for " of that"
 which: use "when" for " at which time"

-
- * Not all the revisions will be appropriate for your document.
 - * When there is more than one suggestion for just one bracketed word, you will have to choose the case that fits your use.

Aug 24 13:01 1982 PROOF OUTPUT FOR kniffin06 Page 4

* Capitalized words are instructions, not suggestions.

NOTE: If you want this program to look for additional phrases or to stop looking for some, for instance to stop flagging "impact," type the command dictadd.

***** SPLIT INFINITIVES *****

For file kniffin06:

No split infinitives found

Aug 24 13:05 1982 SEXIST OUTPUT FOR kniffin06 Page 1

file kniffin06: number of lines 79 number of phrases found 0

ORG OUTPUT FOR kniffin06

SAMPLE 6

Because our endeavors to ameliorate readability and text processing are familiar at numerous echelons of the military, the requirement for an alteration in our standard demonstration therefore became obvious. We anticipate that subsequent to the termination of this demonstration, the human propensity to forget all the technicalities will take effect; however, by retaining the sample transcriptions provided to you, you can facilitate your memory.

Today we are exhibiting one of the primary features of our new computer-aided editing system; this feature has been designated "the substitution dictionary. Each substitute was selected to match the shade of meaning of the difficult word that it might replace, and additionally each substitute is a shorter, more familiar, and more frequently utilized word.

Frequently a writer ascertains from readability analysis that a text is likely to be too difficult for the intended audience. A readability analysis of the edited text can then be accomplished to determine if the text is at an appropriate level of difficulty for the intended audience.

In the near future, we hope to add numerous innovations to the computer-aided editing system. Finally, if we couple together the research on word and phrase substitution, parts of speech analysis, and syntax roles, we will have the necessary requisites for subsequent translation from English to foreign languages.

Aug 24 13:04 1982 PROSE OUTPUT FOR kniffin06 Page 1

BECAUSE YOUR TEXT IS SHORT (< 2000 WORDS & < 100 SENTENCES),
THE FOLLOWING ANALYSIS MAY BE MISLEADING.

NOTE: Your document is being compared against standards
derived from 30 technical memoranda, classified as good
by managers in the research area of Bell Laboratories.

READABILITY

The Kincaid readability formula predicts that your text
can be read by someone with 16 or more years of schooling,
which is rather high for this type of document. Good technical
documents average close to 13th grade level, even
though the audience has more education than that.

This text includes many long words. Consider running
the syllable counting program, syl, to look at the words in
this text with five or more syllables. To do this type the
following command when this program is done.

```
syl -5 filename
```

If most of the long words are technical terms that you must
use, consider providing a glossary of terms to make this
paper easier to read. If the words aren't technical terms,
use shorter words wherever you can. In changing your text,
be sure not to use several different words for the same
concept; such synonyms are frequently confusing to the
reader.

VARIATION

You have an appropriate distribution of sentence types.

SENTENCE STRUCTURE

Passives and Nominalizations

You have appropriately limited your use of passives and
nominalizations (nouns made from verbs, e.g. "description").

PROSE OUTPUTS

Options

You can request that your document be compared against
different standards; typing -t with the prose command, e.g.,

```
prose -t filename
```

will compare your text against training documents.

A -s option will provide a very short version of the
prose output.

Aug 24 13:04 1982 PROSE OUTPUT FOR kniffin06 Page 2

```
prose -s filename
```

If you already have a style table in a file, you can save time by using it as the input to prose rather than the textfile. To do this, precede the style table filename with a -f, e.g.,

```
prose -f styletable-filename
```

All the options can be selected at the same time and listed in any order.

```
prose -f styletable-filename -s -t
```

Statistics

The table of statistics generated by the program style can be found in your file styl.tmp. If you want to look at it type:

```
cat styl.tmp
```

You can also use the match program, which provides a better format, type:

```
match styl.tmp
```

If you are not interested in the file, remove it by typing:

```
rm styl.tmp
```

ORGANIZATION

The prose program cannot check the content or organization of your text. One way to look at the overall structure of your text is to use grep to list all the headings that were specified for the mm formatter. To do this, type:

```
grep '^\.H' filename
```

You can also use the organization program, org, to look at the structure of your text. Org will format your paper with all the headings and paragraph divisions intact, but will only print the first and last sentence of each paragraph in your text so you can check your flow of ideas.

```
org filename
```


Aug 24 13:06 1982 ABST OUTPUT FOR kniffin06 Page 1

Texts differ in the extent to which they refer to concrete objects and abstract ideas. Concrete objects, places, or things can be seen, heard, felt, smelled, or tasted. Abstract ideas, on the other hand, cannot be experienced by our senses. From the results of psychological research, we know that concrete texts are easier to read, easier to use, and easier to remember.

In file "kniffin06," 3.7 percent of the words are abstract words, which is a high score.

Texts with more than about 2% abstract words are abstract. A sample of Plato's Symposium has 4.41% abstract words. It begins, "For we have a custom, and according to our custom any service to another under the idea that he will be improved..." On the other hand, a text with no abstract words contains this passage: "The explosion of a gaseous mixture such as hydrogen and oxygen,..."

One way to improve such text would be to add concrete examples to explain the abstract ideas.

The abstract words in file "kniffin06" are stored in file "ab.kniffin06". Remove it if you wish.

FILE ab.kniffin06

- 3 substitute
- 2 technical
- 2 future
- 2 believe
- 1 standard
- 1 original
- 1 objective
- 1 necessary
- 1 memory
- 1 forget
- 1 compare
- 1 advantage

Aug 24 13:04 1982 SHORT PROSE OUTPUTS WITH DIFFERENT COMPARISONS Page 1

prose -t -s -f sty.kniffin06
BECAUSE YOUR TEXT IS SHORT (< 2000 WORDS & < 100 SENTENCES),
THE FOLLOWING ANALYSIS MAY BE MISLEADING.

Compared to training material.

Reading grade level--16.2: Very high, Good score = >7.8 - 12.4
Variation--Too many compound + compound-complex sentences (27%).
--Sentences are long--avg length = 24.4 words
--Good length = 15.0 to 20.2
>>Shorten sentences by dividing into simple and complex.

Passives--13.0%: Good

Nominalizations--4.0%: High, Good score = 0% - 3.4%

prose -tm -s -f sty.kniffin06
BECAUSE YOUR TEXT IS SHORT (< 2000 WORDS & < 100 SENTENCES),
THE FOLLOWING ANALYSIS MAY BE MISLEADING.

Compared to technical papers.

Reading grade level--16.2: High, Good score = >10.1 - 15.0
Variation--Good sentence type distribution.
Passives--13.0%: Good
Nominalizations--4.0%: Good

APPENDIX 2 TO CHAPTER 12: CRES

----- READABILITY RESULTS -----

Number of Sentences	Number of Words	Number of Syllables
21	466	864
Aver. Number of Words per Sentence	Avg. Number of Syllables per Word	
22.19	1.85	

GRADE LEVEL (Based on DOD Readability Standard)
14.9

----- WORDS NOT ON COMMON WORD LISTS -----

WORD	FREQ	WORD	FREQ
alteration	2	non-technical	1
alterntive	1	ongoing	1
amalgamation	1	practicable	1
ameliorate	1	propensity	1
anaphora	1	readability	4
audience	2	requisites	1
augment	1	roles	1
computer-aided	2	substitution	4
desirable	1	syntax	1
dictionary	1	technicalities	1
elementary	1	transcriptions	1
example	1	translation	1
fabricate	1	vocabulary	1
inference	1	writer	2
innovations	1		

Because our <endeavors><*TRIES/*> to [ameliorate] [readability]
 and text processing are familiar at <numerous><*MANY/MOST*>
 <echelons><*LEVELS/*> of military, the <requirement><*NEED/*>
 for an [alteration] in our standard demonstration
 <therefore><*SO/*> became obvious.<*031*> <Accordingly,><*SO/*>
 you are <observing><*SEEING/OBEYING*> an [amalgamation] of the
 products of our [ongoing] and future research with our
 <objective><*AIM/GOAL*> being to <exhibit><*SHOW/*> a
 <consolidated><*UNITED/MADE FIRM*> demonstration of our efforts
 at producing <beneficial><*HELPPFUL/*>
 <modifications><*CHANGES/*> in both text and text
 processing.<*037*><* TOO MANY (07) PREPOSITIONS *> We
 <anticipate><*EXPECT/*> that <subsequent to><*AFTER/LATER*> the
 <termination of><*ENDING/FINISHING*> this demonstration, the
 human [propensity] to forget all the [technicalities] will take
 <effect;><*MAKE/BRING ABOUT*> however, by
 <retaining><*KEEPING/*> the sample [transcriptions]
 <provided><*GAVE/GIVEN/SUPPLIED*> to you, you can
 <facilitate><*EASE/HELP*> your memory. Today we are
 <exhibiting><*SHOWING/*> one of the primary features of our new
 [computer-aided] editing system; this feature has been
 <designated><*CERTAIN/CHOSEN*> "the [substitution]
 [dictionary]." One purpose of a [substitution] [dictionary] is
 to [augment] [readability] by suggesting [alternative]
 [elementary] words to replace the difficult [non-technical]
 words in a text.<*024> The [dictionary]
 <incorporates><*BLENDS/JOINS*> two word lists. One list

<comprises><*FORMS/INCLUDES*> <approximately><*ABOUT/*> 900
 difficult words that frequently <appear><*SEEM/*> in technical
 manuals while the second list <comprises><*FORMS/INCLUDES*>
 <approximately><*ABOUT/*> 1800 suggested substitutes for the
 words in the original list.<*029*> Each substitute was
 <selected><*CERTAIN/*> to match the shade of meaning of the
 difficult word that it might replace, and additionally each
 substitute is a shorter, more familiar, and more frequently
 <utilized><(USED/*> word.<*032*> Frequently a [writer]
 <ascertains><*FINDS OUT/LEARNS*> from [readability] analysis
 that a text is likely to be too difficult for the intended
 [audience.] The text processor is then <utilized><*USED/*> to
 compare each word in the text with the list of difficult words
 in the [substitution] [dictionary.]<*023*><* TOO MANY (05)
 PREPOSITIONS *> When a word in the text corresponds to a word in
 the difficult word list, the text processor prints out the
 corresponding substitute words.<*024*> The [writer] must then
 <determine><*DECIDE/FIGURE*> which word
 <modifications><*CHANGES/*> can then be made<* PASSIVE VOICE *>
 without producing an [alteration] in the meaning of the text. A
 [readability] analysis of the edited text can then be
 <accomplished><*CARRIED OUT/DID/DONE*> to
 <determine><*DECIDE/FIGURE*> if the text is at an
 <appropriate><*PROPER/RIGHT*> level of difficulty for the
 intended [audience.]<*027*><*TOO MANY (05) PREPOSITIONS *> In
 the near future, we hope to add <numerous><*MANY/MOST*>
 [innovations] to the [computer-aided] editing system.<* TOO MANY

(04) PREPOSITIONS *> At this time, we believe that [desireable] <benefits><*HELPS/IS HELPED*> will result by means of taking advantage of characteristics that are <peculiar to><*RELATING/*> certain structures.<*023*> Other <benefits><*HELPS/IS HELPED*> could follow after we conduct an investigation of the manner in which writers make references to other parts of their own writing.<*024*><*TOO MANY (05) PREPOSITIONS *> For [example,] we believe that the identification of passive voice, [anaphora,] and [interference] loading may be [practicable.] Futhermore, we are planning to <implement><*CARRY OUT/DO*> the U.S. Navy's 4000 word controlled [vocabulary] and to [fabricate] our own "Controlled Logistics Vocabulary." Finally, if<*SUBORINATE CLAUSE *> we couple together the research on word and phase [substitution,]<* CLAUSE TOO LONG *> parts of speech analysis, and [syntax] [roles,] we will have the necessary [requisites] for <subsequent><*LATER/NEXT*> [translation] from English to foreign languages.<*033*>

 NAME OF SAMPLE TEXT: SAMPLE 6 17 SEP 82

 File begins with:

Because our endeavors to ameliorate readability and text processing are familiar at numerous echelons of the military, the

 LONG WORDS:

endeavors	ameliorate	readability
processing	familiar	numerous
echelons	military	requirement
alteration	demonstration	therefore
obvious	Accordingly	observing
amalgamation	objective	exhibit
consolidated	demonstration	producing
beneficial	modifications	processing
anticipate	subsequent	termination
demonstration	propensity	technicalities
however	retaining	transcriptions
provided	facilitate	memory
exhibiting	primary	computer-aided

producing	alteration	readability
analysis	edited	accomplished
determine	appropriate	difficulty
intended	numerous	innovations
computer-aided	editing	desirable
benefits	advantage	characteristics
peculiar	benefits	investigation
references	example	identification
anaphora	inference	practicable
Furthermore	implement	vocabulary
fabricate	Logistics	Vocabulary
Finally	together	substitution
analysis	necessary	requisites
subsequent	translation	languages

 SUMMARY AND CALCULATIONS:

NUMBER OF SENTENCES= 21
 NUMBER OF WORDS= 463
 NUMBER OF SYLLABLES= 869
 AVERAGE SENTENCE LENGTH= 22.0
 AVG SYLLABLES PER WORD= 1.88
 FL-KIN GRADE LEVEL EQUIV= >15

 SUMMARY AND CALCULATIONS:

NUMBER OF SENTENCES= 21
 NUMBER OF WORDS= 463
 NUMBER OF SYLLABLES= 869
 AVERAGE SENTENCE LENGTH= 22.0
 AVG SYLLABLES PER WORD= 1.88
 FLESCH INDEX= 25.7
 GRADE LEVEL EQUIV= 16.6

NAME OF SAMPLE TEXT: SAMPLE 6

Because our endeavors to ameliorate readability and text
 ---endeavor try, attempt
 ---ameliorate improve, better

processing are familiar at numerous echelons of the military, the
 ---familiar well known
 ---numerous many, some
 ---echelon level

requirement for an alteration in our standard demonstration therefore
 ---requirement need
 ---alteration change
 ---demonstration show
 ---therefore so

became obvious. Accordingly, you are observing an amalgamation of
 ---obvious plain, clear
 -SENTENCE GREATER THAN 20 WORDS.-
 ---Accordingly therefore, so, then
 ---observing noting, seeing
 ---amalgamation combination, blend, mixture

the products of our ongoing and future research with our objective
 ---ongoing present
 ---objective aim, goal

being to exhibit a consolidated demonstration of our efforts at pro-
 ---exhibit show, present
 ---consolidate unite, combine, join
 ---demonstration show

ducing beneficial modifications in both text and text processing. We
 ---producing making
 ---beneficial helpful
 ---modification change
 -SENTENCE GREATER THAN 20 WORDS.-

anticipate that subsequent to the termination of this demonstration,
 ---anticipate expect
 ---subsequent next, later, after
 ---termination end, conclusion
 ---demonstration show

the human propensity to forget all the technicalities will take
 ---propensity tendency
 ---technicalities details

effect; however, by retaining the sample transcriptions provided to

****-SENTENCE GREATER THAN 20 WORDS.-****

---however but, though

---retaining keeping, holding

---transcription copy

---provided given

you, you can facilitate your memory.

---facilitate make easy, help

Today we are exhibiting one of the primary features of our new

---exhibiting showing, presenting

---primary major, first, main, basic

computer-aided editing system; this feature has been designated "the

---designate name, appoint, choose, pick, select

substitution dictionary." One purpose of a substitution dictionary

is to augment readability by suggesting alternative elementary words

---augment increase

---alternative choice, choice of, option

---elementary basic, simple

to replace the difficult non-technical words in a text. The diction-

---difficult hard

****-SENTENCE GREATER THAN 20 WORDS.-****

ary incorporates two word lists. One list comprises approximately

---incorporate combine, include, join

---comprise consist, contain, include, make up

---approximately about, almost

900 difficult words that frequently appear in technical manuals while

---difficult hard

---frequently often

the second list comprises approximately 1800 suggested substitutes

---comprise consist, contain, include, make up

---approximately about, almost

for the words in the original list. Each substitute was selected to

---original first, new

****-SENTENCE GREATER THAN 20 WORDS.-****

match the shade of meaning of the difficult word that it might

---difficult hard

replace, and additionally each substitute is a shorter, more

---additionally too, also

familiar, and more frequently utilized word.

---familiar well known

---frequently often

---utilize use

****-SENTENCE GREATER THAN 20 WORDS.-****

Frequently a writer ascertains from readability analysis that a
 ---Frequently often
 ---ascertain find out, learn, make sure

text is likely to be too difficult for the intended audience. The
 ---difficult hard

text processor is then utilized to compare each word in the text with
 ---utilize use

the list of difficult words in the substitution dictionary. When a
 ---difficult hard

-SENTENCE GREATER THAN 20 WORDS.-

word in the text corresponds to a word in the difficult word list,
 ---correspond agree, match
 ---difficult hard

the text processor prints out the corresponding substitute words.
 ---corresponding agreeing, matching
 -SENTENCE GREATER THAN 20 WORDS.-

The writer must then determine which word modifications can be made
 ---determine decide, find out
 ---modification change

without producing an alteration in the meaning of the text. A read-
 ---producing making
 ---alteration change
 -SENTENCE GREATER THAN 20 WORDS.-

ability analysis of the edited text can then be accomplished to
 ---accomplished expert, done

determine if the text is at an appropriate level of difficulty for
 ---determine decide, find out
 ---appropriate proper, set aside
 ---difficult hard

the intended audience.
 -SENTENCE GREATER THAN 20 WORDS.-

In the near future, we hope to add numerous innovations to the
 ---numerous many, some
 ---innovation change

computer-aided editing system. At this time, we believe that desir-
 able benefits will result by means of taking advantage of character-
 ---benefit help

istics that are peculiar to certain structures. Other benefits could
 ---characteristic feature, quality, trait
 ---peculiar odd, strange, special
 -SENTENCE GREATER THAN 20 WORDS.-
 ---benefit help

follow after we conduct an investigation of the manner in which

writers make references to other parts of their own writing. For

****--SENTENCE GREATER THAN 20 WORDS.--****

example, we believe that the identification of passive voice,

anaphora, and inference loading may be practicable. Furthermore, we

---practicable can be done, possible

---Furthermore also, too, besides

are planning to implement the U.S. Navy's 4000 word controlled vocab-

---implement follow, effect, tool

ulary and to fabricate our own "Controlled Logistics Vocabulary."

---fabricate build, make

Finally, if we couple together the research on word and phrase

substitution, parts of speech analysis, and syntax roles, we will

have the necessary requisites for subsequent translation from English

---necessary needed

---requisite needed

---subsequent next, later, after

to foreign languages.

****--SENTENCE GREATER THAN 20 WORDS.--****

NAME OF SAMPLE TEXT: SAMPLE 6

NEG/NOM/ABSTRACT WORD CHECK

Because our endeavors to ameliorate readability and text processing are familiar at numerous echelons of the military the requirement for an alteration in our standard demonstration therefore became obvious Accordingly you are observing an amalgamation of the products of our ongoing and future research with our objective being to exhibit a consolidated demonstration of our efforts at producing beneficial modifications in both text and text processing We anticipate that subsequent to the termination of this demonstration the human propensity to forget all the technicalities will take effect however by retaining the sample transcriptions provided to you you can facilitate your memory Today we are exhibiting one of the primary features of our new computer-aided editing system this feature has been designated the substitution dictionary One purpose of a substitution dictionary is to augment readability by suggesting alternative elementary words to replace the difficult non-technical words in a text The dictionary incorporates two word lists One list comprises approximately 900 difficult words that frequently appear in technical manuals while the second list comprises approximately 1800 suggested substitutes for the words in the original list Each substitute was selected to match the shade of meaning of the difficult word that it might replace and additionally each substitute is a shorter more familiar and more frequently utilized word Frequently a writer ascertains from readability analysis that a text is likely to be too difficult for the intended audience The text processor is then utilized to compare each word in the text with the list of difficult words in the substitution dictionary When a word in the text corresponds to a word in the difficult word list the text processor prints out the corresponding substitute words

The writer must then determine which word modifications can be made without
 'NOM
 producing an alteration in the meaning of the text A readability analysis
 'NOM
 of the edited text can then be accomplished to determine if the text is at
 an appropriate level of difficulty for the intended audience In the near
 future we hope to add numerous innovations to the computer-aided editing
 'NOM
 system At this time we believe that desirable benefits will result by means
 of taking advantage of characteristics that are peculiar to certain structures
 'ABS
 Other benefits could follow after we conduct an investigation of the manner
 'NOM'ABS
 in which writers make references to other parts of their own writing For
 'NOM
 example we believe that the identification of passive voice anaphora and
 'NOM
 inference loading may be practicable Furthermore we are planning to implement
 'NOM
 the U.S Navy's 4000 word controlled vocabulary and to fabricate our own Controlled
 Logistics Vocabulary Finally if we couple together the research on word and
 phrase substitution parts of speech analysis and syntax roles we will have
 'NOM
 the necessary requisites for subsequent translation from English to foreign
 'NOM
 languages